



Basi Di Dati e di conoscenza

Calcolo Relazionale



Contenuti della lezione

- Calcolo relazionale
 - Calcolo sui domini
 - Calcolo sulle ennuple
- Espressioni in calcolo relazionale
- Equivalenza fra i linguaggi

Operazioni Modello relazionale

- Notazione algebrica \Rightarrow **Algebra relazionale** \Rightarrow Linguaggio procedurale \Rightarrow interrogazioni espresse applicando operatori alle relazioni
- Notazione logica \Rightarrow **Calcolo relazionale** \Rightarrow Linguaggio dichiarativo \Rightarrow interrogazioni espresse tramite formule logiche le cui risposte devono essere rese vere dalle tuple

Calcolo relazionale

- Basato sulla logica del prim'ordine
 - Linguaggio formale per la rappresentazione della conoscenza
 - Semantica non ambigua
 - Sistema formale di inferenza
- Calcolo relazionale è alla base di quasi tutti i linguaggi di interrogazione esistenti e basati sul modello relazionale.
- Esistono diverse versioni del calcolo relazionale, ne vedremo solo due :
 - **Calcolo relazionale sui domini** (più vicino al calcolo dei predicati)
 - **Calcolo relazionale sulle tuple con dichiarazioni di range**
 - Variazione del precedente
 - Base di molti costrutti degli attuali linguaggi

Calcolo relazionale

- Il calcolo dei domini è una 6-pla
 - $\{A, D, \text{dom}, s, O, F\}$
 - A: Insieme degli attributi
 - D: Insieme dei domini
 - Dom: $A \rightarrow D$
 - s: Schema di base di dati
 - O: Insieme degli **operatori di confronto** ($>, \geq, <, \leq, \neq, =$) e **logici** (\wedge, \vee, \neg) e i quantificatori esistenziale (\exists) e universale (\forall).
 - F è l'insieme delle formule corrette

Formule corrette:

- Una formula corretta è definita ricorsivamente a partire dagli atomi che sono formule corrette:
 - Atomi:
 - $R(x)$ dove R appartenente a s è uno schema di relazione e x è una *variabile di ennupla* (**Calcolo delle Ennuple**)
oppure
 - $R(A_1: x_1, \dots, A_p: x_p)$ dove $R(A_1, \dots, A_p)$ è uno schema di relazione appartenente a s e x_1, \dots, x_p sono variabili di dominio (**Calcolo dei domini**)
 - $x\theta y$ o $x\theta c$, con x e y variabili di ennupla (risp. variabili di dominio), c è una costante e θ operatore di confronto
 - Se f_1 e f_2 sono formule corrette, allora $f_1 \wedge f_2$, $f_1 \vee f_2$, $\neg f_1$, (f_1) formule corrette (le parentesi sono utilizzate per alterare il normale ordine di precedenza nelle espressioni (\neg, \wedge, \vee)).
 - Se f è una formula corretta e x è una variabile di ennupla (risp di dominio), allora $\exists x(f)$ e $\forall x(f)$ sono formule corrette.

Espressioni nel calcolo relazionale

- Una espressione nel calcolo relazionale (query) ha la seguente forma:
 - **Calcolo dei domini** : $\{A_1: x_1, \dots, A_p: x_p \mid f\}$
 - $A_1: x_1, \dots, A_p: x_p$ è la target list
 - Dove A_1, \dots, A_p sono attributi distinti e
 - x_1, \dots, x_p sono variabili di dominio che rendono vera la formula corretta f
 - **Calcolo delle ennuple**
 - $\{x \mid f\}$
 - Dove x è una variabile di ennupla che rende vera la formula corretta f

Verità delle formule

- Una formula atomica :
 - $R(x)$ è **vera** sui valori di x che rappresentano ennuple di R (**Calcolo delle Ennuple**), oppure
 - $R(A_1: x_1, \dots, A_p: x_p)$ è vera sui valori x_1, \dots, x_p che formano una ennupla di R (**Calcolo dei domini**)
 - $x\theta y$ o $x\theta c$, è vera sui valori a_1 e a_2 tale che $a_1\theta a_2$ o $a_1\theta c$ sono soddisfatte:
- La verità delle formule costruite per congiunzione, disgiunzione e negazione valgono le regole usuali,
- Le formule con i quantificatori sono vere secondo le seguenti regole:
 - $\exists x(f)$ è vera se esiste **almeno** un valore a per la variabile x che rende vera la formula f
 - $\forall x(f)$ è vera se per **ogni** possibile valore a per la variabile x , la formula f risulta vera.

Calcolo relazionale- Interrogazioni

Impiegati

Matricola	Cognome	Età	Stipendio
101	Rossi	34	40
103	Bianchi	23	35
104	Neri	38	61
210	Celli	49	60
231	Bisi	50	60
252	Bini	44	70
301	S. Rossi	34	70
375	M. Rossi	50	65

Supervisione

Capo	Impiegato
210	101
210	103
210	104
301	210
301	231
375	252

Calcolo relazionale- Interrogazioni

- matricola, nome ed età degli impiegati che guadagnano più di 40mila euro

$\{\text{Matricola:m, Cognome:n, Et\grave{a}:e} \mid \text{Impiegati}(\text{Matricola :m, Cognome :n, Et\grave{a}:e, Stip:s}) \wedge s > 40\}$

- nome degli impiegati che guadagnano più di 40mila euro

$\{\text{Cognome:n} \mid (\exists m) (\exists e) \text{ Impiegati}(\text{Matricola:m, Cognome :n, Et\grave{a}:e, Stip:s}) \wedge s > 40\}$

Matr icola	Cognome	Et\grave{a}
104	Neri	38
210	Celli	49
231	Bisi	50
252	Bini	44
301	S. Rossi	34
375	M. Rossi	50

Cognome
Neri
Celli
Bisi
Bini
S. Rossi
M. Rossi

Problemi con il calcolo relazionale: Assunzione di mondo chiuso

- Il calcolo relazionale ammette espressioni senza senso (sintatticamente corrette e semanticamente non valide)
 - $\{A_1: x_1, A_2: x_2 \mid R(A_1: x_1) \wedge (x_2 = x_2)\}$
 - $\{A_1: x_1 \mid \neg (R(A_1: x_1))\}$
 - Il risultato cambia al cambiare del dominio e può essere infinito se il dominio è infinito
- Un linguaggio di interrogazione è **indipendente dal dominio** se il suo risultato, su ciascuna istanza di base di dati, non varia al variare del dominio rispetto al quale l'espressione è valutata.
- Si assume l'ipotesi di **mondo chiuso** in cui i domini sono ristretti ai valori presenti nell'istanza dello schema relazionale e alle costanti presenti nelle espressioni.
- Sotto questa ipotesi il calcolo relazionale è un linguaggio indipendente dal dominio.

Calcolo relazionale

Considerazioni:

- un' espressione di un linguaggio di interrogazione sarebbe utile che fosse indipendente dal dominio
- Abbiamo bisogno di un' altra versione del calcolo relazionale, in cui le variabili, anziché denotare singoli valori, denotino tuple.

Calcolo relazionale sui Domini ha dei difetti :

- Agisce sui domini invece che sui valori
- Per il motivo precedente diventa "verboso" (ha bisogno di tante variabili)
- Può portare a espressioni che non hanno senso
- Occorre un linguaggio che 'focalizzi' le tuple di interesse

Calcolo relazionale su tuple

- **Calcolo relazioni su tuple:**

Espressione: {Target list | Range list | formula }

- **Target list:** lista degli obiettivi con elementi $Y:x.Z$ o $x.Z$ se $Z:x.Z$ o $x.*$
- **Range list:** elenco delle variabili libere della formula con i relativi campi di variabilità
- **Formula è del tipo:**
 - $x.A\theta c$ o $x.A\theta y.B$
 - connettivi di formule
 - $\exists x(R)(f)$ o $\forall x(R)(f)$

Calcolo relazionale- Interrogazioni

- matricola, nome ed età degli impiegati che guadagnano più di 40mila euro

{i.(Matr, Nome, Età) | i(IMPIEGATI) | i.Stip > 40}

Matr icola	Cognome	Età
104	Neri	38
210	Celli	49
231	Bisi	50
252	Bini	44
301	S. Rossi	34
375	M. Rossi	50

{i.* | i(IMPIEGATI) | i.Stip > 40} con * prendo tutti gli attributi

Calcolo relazionale

Considerazioni:

- Il calcolo su tuple però non permette di esprimere tutte le interrogazioni che possono essere formulate in Algebra relazionale.
- **Esempio:** non c'è l'unione, per questo nei linguaggi interrogativi viene aggiunto esplicitamente un costrutto di unione.

Calcolo relazionale- Interrogazioni

- Trovare gli impiegati che guadagnano più del proprio capo, mostrando, nome e stipendio dell'impiegato e del capo

- **Algebra relazionale**

$$\pi_{\text{Nome,Stip,MatrC,NomeC,StipC}} (\sigma_{\text{Stipendio} > \text{StipC}} ($$
$$\rho_{\text{MatrC,NomeC,StipC,EtàC}} \leftarrow \text{Matr,Nome,Stip,Età} (\text{Impiegati}))$$
$$\bowtie_{\text{MatrC=Capo}} (\text{Supervisione} \bowtie_{\text{Impiegato=Matricola}} \text{Impiegati)))$$

- **Calcolo dei domini**

$$\{ \text{Nome } n, \text{Stip } s, \text{NomeC } nc, \text{StipC } sc \mid \text{Impiegati}(\text{Matr: } m, \text{Nome } n, \text{Età } e, \text{Stipendio: } s) \wedge (s > sc) \wedge \text{Supervisione}(\text{Impiegato: } m, \text{Capo: } c) \wedge \text{Impiegati}(\text{Matr: } c, \text{Nome } nc, \text{Età } ec, \text{Stipendio: } sc) \}$$

- **Calcolo delle ennuple**

$$\{ t^{[4]} \mid (\exists x) (\exists y) (\exists z) \text{Impiegati}(x) \wedge \text{Supervisione}(y) \wedge (y.\text{Impiegato} = x.\text{matr}) \wedge \text{Impiegati}(z) \wedge (y.\text{Capo} = z.\text{Matr}) \wedge (t.\text{Nome} = x.\text{Nome}) \wedge (t.\text{Stip} > x.\text{Stip}) \wedge (t.\text{NomeCapo} = z.\text{Nome}) \wedge (t.\text{StipCapo} = z.\text{Stip}) \}$$

Calcolo relazionale- Interrogazioni

- Trovare le matricole e i nomi dei capi i cui impiegati guadagnano **tutti** più di 40 milioni

- Algebra relazionale

$$\pi_{\text{Capo}}(\text{Supervisione}) - \pi_{\text{Capo}}(\text{Supervisione} \bowtie_{\text{Impiegato=Matricola}} (\sigma_{\text{Stipendio} \leq 40}(\text{Impiegati})))$$

- Calcolo dei domini (due modi: negazione quantificatore esistenziale)

- $\{\text{Matricola: } c, \text{ Nome } n \mid \text{Impiegati}(\text{Matr: } c, \text{ Nome } n, \text{ Et\`a } e, \text{ Stipendio: } s) \wedge \text{Supervisione}(\text{Impiegato: } m, \text{ Capo: } c) \wedge \neg \exists m' (\exists n' (\exists e' (\exists s' (\text{Impiegati}(\text{Matr: } m', \text{ Nome } n', \text{ Et\`a } e', \text{ Stipendio: } s') \wedge \text{Supervisione}(\text{Impiegato: } m', \text{ Capo: } c) \wedge (s' \leq 40)))\}$
- $\{\text{Matricola: } c, \text{ Nome } n \mid \text{Impiegati}(\text{Matr: } c, \text{ Nome } n, \text{ Et\`a } e, \text{ Stipendio: } s) \wedge \forall m' (\forall n' (\forall e' (\forall s' (\text{Impiegati}(\text{Matr: } m', \text{ Nome } n', \text{ Et\`a } e', \text{ Stipendio: } s') \wedge \text{Supervisione}(\text{Impiegato: } m', \text{ Capo: } c) \wedge (s' > 40)))\}$

Equivalenza fra i linguaggi

- E' possibile dimostrare che:
 - Per ogni espressione del calcolo relazionale che sia indipendente dal dominio esiste un'espressione dell'algebra relazionale equivalente ad essa;
 - Per ogni espressione dell'algebra relazionale esiste un'espressione del calcolo relazionale equivalente ad essa.

Dim: In modo ricorsivo a partire dagli operatori di base.

SQL



